DOCUMENT RESUME

ED 111 758 SP 009 302

AUTHOR Trzasko, Joseph A.
TITLE Statistical Considerations for Establishing CBTE Cut-Off Scores.
PUB DATE 75
NOTE 14p.
AVAILABLE FROM Joseph A. Trzasko, Ph.D., Department of Psychology, Mercy College, Dobbs Ferry, New York 10522 (No price quoted)

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS *Cutting Scores; *Decision Making; *Performance Based Teacher Education; Performance Criteria; Reliability; Statistical Analysis; *Test Construction

ABSTRACT

This report gives the basic definition and purpose of competency-based teacher education (CBTE) cut-off scores. It describes the basic characteristics of CBTE as a yes-no dichotomous decision regarding the presence of a specific ability or knowledge, which necesitates the establishment of a cut-off point to designate competency vs. incompetency on stated objectives. Statistical considerations for establishing CBTE cut-off scores are reviewed, and, based on test scores, two types of classification errors are identified. These are false acceptance, i.e., nonmasters erroneously classified as masters; and false rejection, i.e., masters erroneously classified as nonmasters. The report recommends that cut-off scores should not be arbitrarily established, but should be based on decision theory, the goal of which is to increase correct rate and decrease error rate. It was found that increasing the cut-off score will decrease false acceptance error and increase false rejection error, while decreasing the cut-off score will decrease false rejection error and increase false acceptance error. It was concluded that since shifting cut-off points reduces one error at the expense of the other, the direction of shift should be in the direction of the less serious error (false rejection). Numerous figures and tables are included. (BD)

STATISTICAL CONSIDERATIONS FOR ESTABLISHING CBTE CUT-OFF SCORES

Joseph A. Trzasko, Ph.D.
Dept. of Psychology
Mercy College

The basic definition and purpose of cut-off scores is discussed with regard to competency-based teacher education (CBTE). Statistical considerations for establishing CBTE cut-off scores are reviewed, including the concepts of correct rejection, false acceptance, false rejection, and correct acceptance. The basic application of decision theory to CBTE is offered.

A basic characteristic of competency-based teacher education (CBTE) is its yes-no dichotomous decision regarding the presence of a specific ability or knowledge, i.e., does a student display competency or incompetency regarding required performance or academic material. Therefore, a cut-off point must be established to designate competency vs. incompetency on stated objectives. Beggs & Lewis (1975, pg. 61-62) summarize the use of cut-off scores as follows:

> <u>Cut-off scores</u>. One term most frequently linked with criterion-referenced measurement is <u>cut-off score</u>. The connection between them occurs because they are both used in situations in which the concern is to determine whether a student possesses certain behavior. A cut-off score is generally applied when a teacher is teaching for mastery, attempting to cause students to reach a point at which they can answer some percentage of items on a test correctly. The cut-off score is the score the student must obtain before the teacher is willing to accept that the student has mastered the topic or content under consideration. The cut-off score is further interpreted to indicate the minimal level of the skill being evaluated that the student must possess to be successful at the next level. Cut-off scores are therefore generally established at the upper end of a scale. That is, a student may be required to respond correctly to 85, 90, or 95 percent of the items. Cut-off scores and criterion-referenced measurement are frequently confused because it is popular among teachers, evaluators, and others to call the cut-off score the "criterion score."
>
> Nothing in criterion-referenced measurement requires the use of a cut-off score. Criterion-referenced measurement is designed to determine whether an individual possesses certain skills. The criterion in

criterion-referenced measurement is the test item. A correct response indicates that the individual possesses that skill. An incorrect response indicates that the skill has not been achieved. Cut-off scores, on the other hand, indicate that the student must achieve some minimal percentage on a test. The score that the student obtains is interpreted to mean that the student has mastered that percentage of the content.

Cut-off scores should not be arbitrarily established, but should be based on decision theory. The end-product of the correct use of decision theory will be a reduction in the number of erroneous judgments. For example, assume that a given test is administered to two groups: one group known to have mastered the material and another group known not to have mastered the material. And, as a result of within-group variability, the two frequency distributions overlap. This is illustrated in Figure 1. The cut-off point is

---

Insert Figure 1

---

set at 70; therefore, according to the test, students scoring 70 and above are considered masters, while those students scoring below 70 are considered non-masters. It will be noted that based on test scores, two types of classification errors will have been made: (1) false acceptance, i.e., non-masters erroneously classified as masters; and, (2) false rejection, i.e., masters erroneously classified as non-masters. The overall number of correct decisions or 'correct rate' is the frequency of correct rejection plus the frequency of correct acceptance, while the overall number of incorrect decisions or 'error rate' is the frequency of false acceptance plus the frequency of false rejection. <u>The goal of decision theory is to increase correct rate and decrease error rate.</u>

Increasing the cut-off score, e.g., to 75, will simultaneously decrease the false acceptance error and increase the false rejection error. And, decreasing the cut-off score, e.g., to 65, will simultaneously decrease the false rejection error and increase the false acceptance error. Increasing

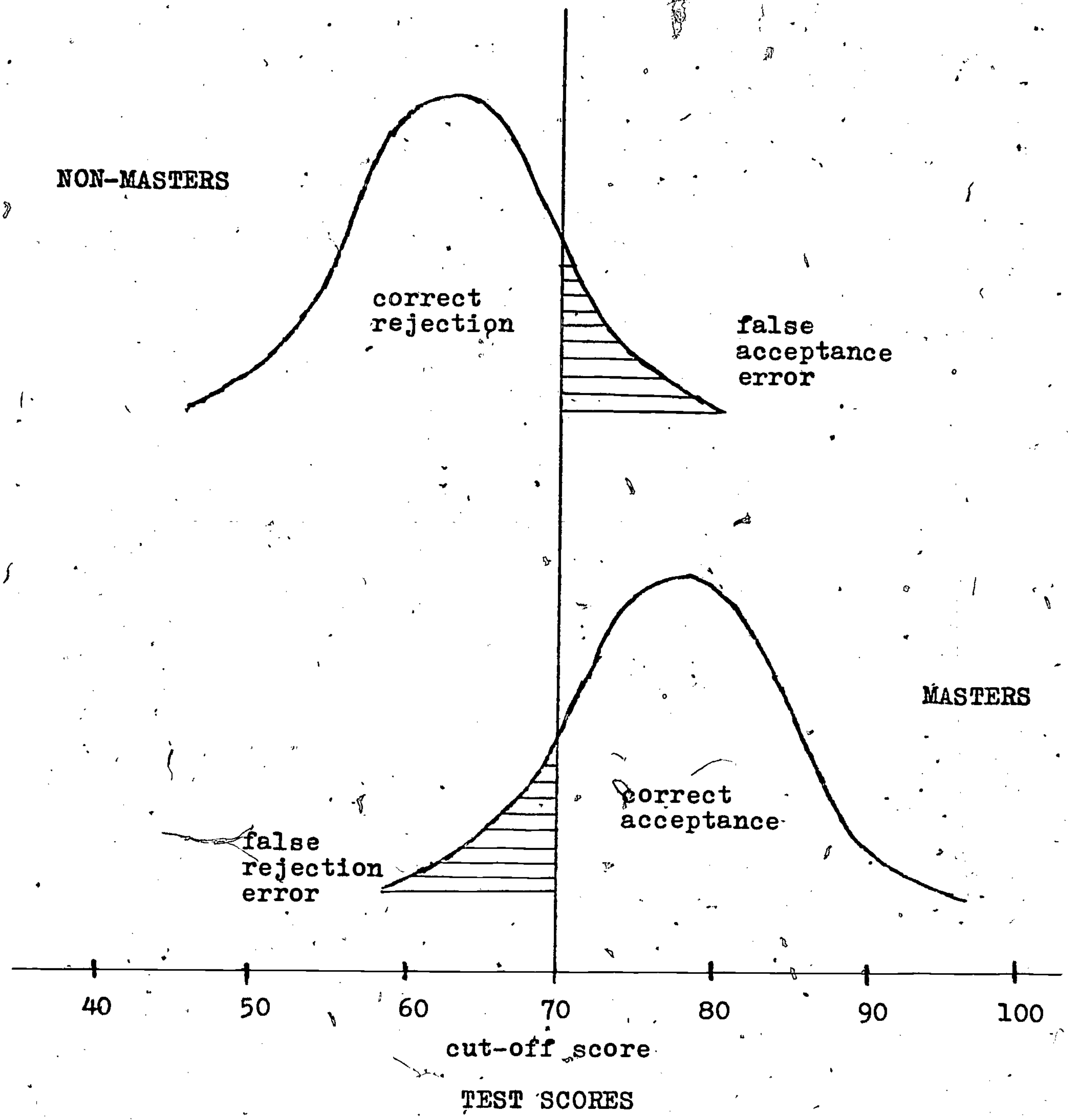

Fig. 1. Frequency distributions of test scores for masters and non-masters with cut-off score of 70 indicated.

and decreasing the cut-off point is illustrated in Figures 2 and 3, respectively. Therefore, simply increasing or

Insert Figures 2 & 3

decreasing the cut-off point will decrease one type of error at the expense of the other. One would then have to decide whether increasing false acceptance or false rejection is more serious.

Figures 1 through 3 can be represented in decision matrices. The decision matrices for the various cut-off points are presented in Table 1. It should be noted that,

Insert Table 1

in the examples given, the number of masters and non-masters were equal, i.e., $P(M) = P(NM) = .50$ with regard to the total number of students used. The vertical column totals reflect the equality or inequality of group size. The cell entries can best be described as joint probabilities, as presented for cut-off score of 65:

$$P(M_{TS} \cap M_J) = P(M_{TS}) \times P(M_J | M_{TS})$$
$$.50 \quad \times \quad .90 \quad = \quad .45$$
$$P(M_{TS} \cap NM_J) = P(M_{TS}) \times P(NM_J | M_{TS})$$
$$.50 \quad \times \quad .10 \quad = \quad .05$$
$$P(NM_{TS} \cap M_J) = P(NM_{TS}) \times P(M_J | NM_{TS})$$
$$.50 \quad \times \quad .40 \quad = \quad .20$$
$$P(NM_{TS} \cap NM_J) = P(NM_{TS}) \times P(NM_J | NM_{TS})$$
$$.50 \quad \times \quad .60 \quad = \quad .30$$

The probability of correct judgment = .75, i.e., $P(M_{TS} \cap M_J) + P(NM_{TS} \cap NM_J)$; and, the probability of incorrect judgment = .25, i.e., $P(M_{TS} \cap NM_J) + P(NM_{TS} \cap M_J)$. It is then possible to compare correct and error rates per each cut-off point.

As indicated above, one type of error may be judged as more serious, i.e., the experimenter may be able to subjec-

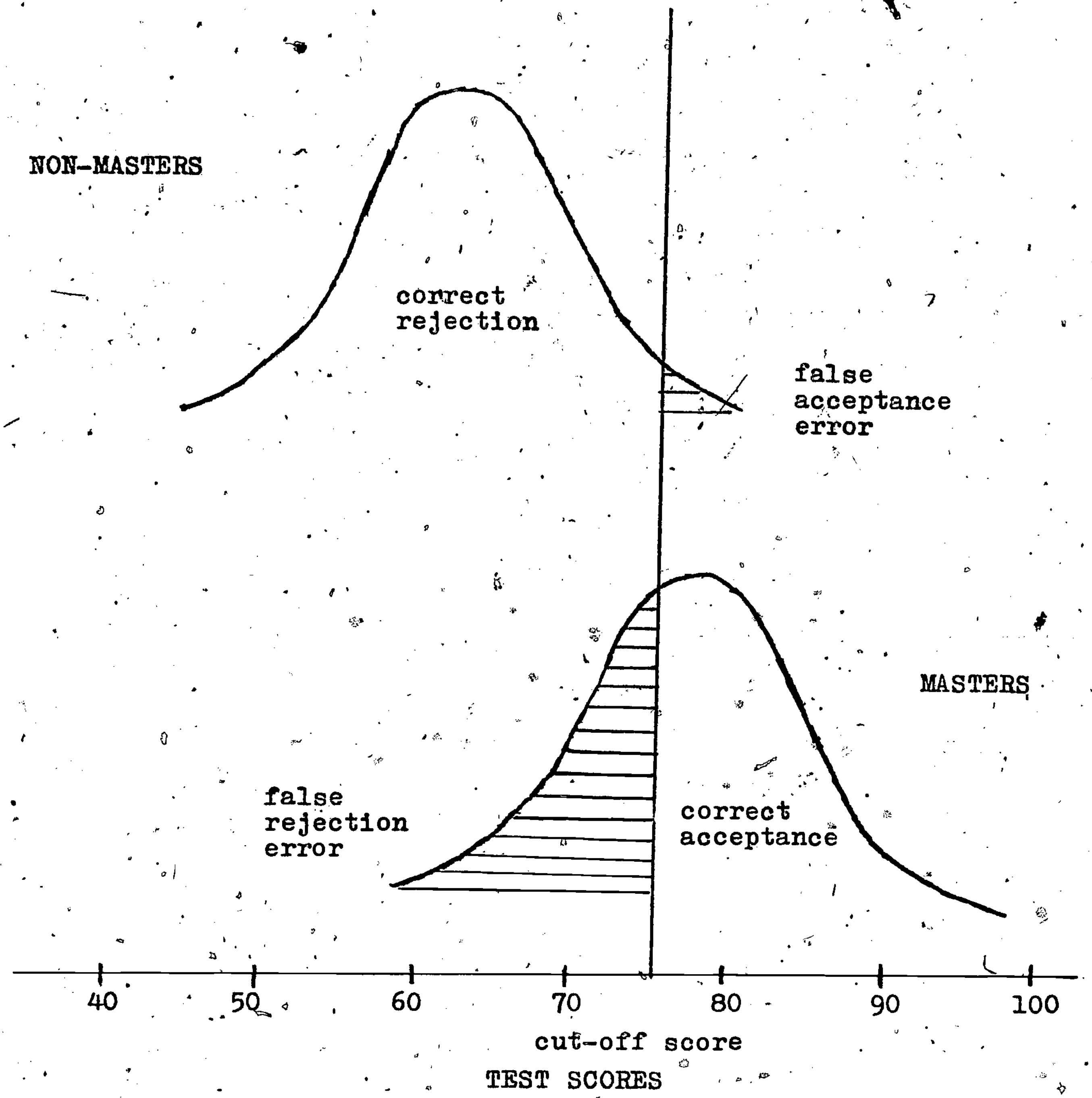

Fig. 2. Frequency distributions of test scores for masters and masters with cut-off score increased to 75.

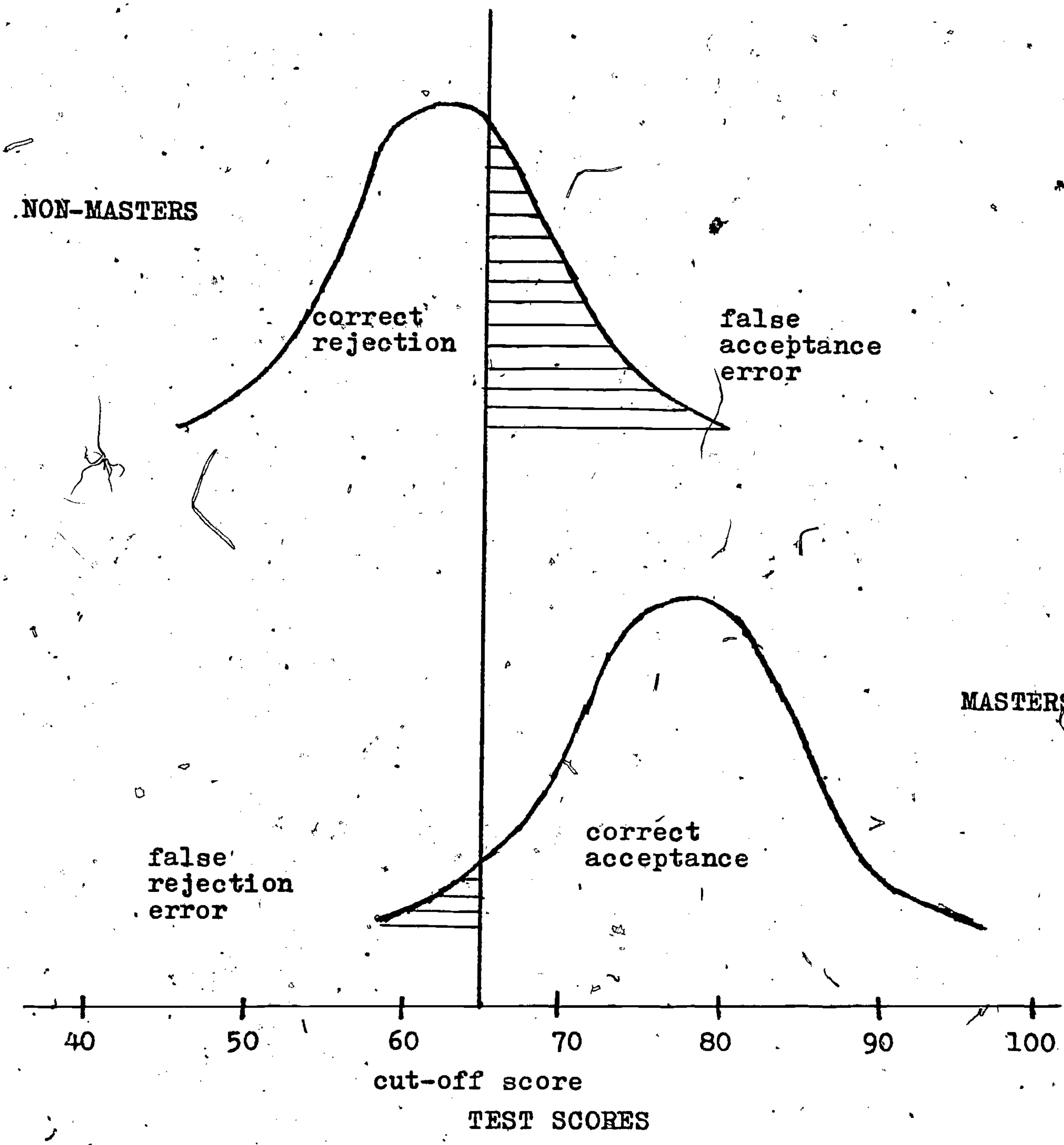

Fig. 3. Frequency distributions of test scores for masters and non-masters with cut-off score decreased to 65.

TABLE I

DECISION MATRICES FOR THE VARIOUS CUT-OFF POINTS

		TRUE STATE	
		Master	Non-Master
JUDGMENT	Master	correct	false acceptance
	Non-Master	false rejection	correct

CUT-OFF SCORE: 65

.45	.20
.05	.30

CUT-OFF SCORE: 70

.35	.15
.[illegible]5	.35

CUT-OFF SCORE: 75

.30	.05
.20	.45

tively or objectively quantify the relative weight of the two types of error. For example, while no error (0) is entailed in correct rejection or correct acceptance, the relative seriousness of false acceptance may be judged twice that of false rejection. Therefore, false acceptance would be assigned a relative weight of 2 and false rejection would be assigned a relative weight of 1. That is, as based on available data from students and faculty, it may be judged twice as serious to misclassify a non-master than a master. While the misclassified master would have to repeat the material, the misclassified non-master, erroneously judged as competent, may not be exposed to the material. In the educational field, erroneously labeling incompetent student-teachers as competent could have a significant effect on pupil performance, while erroneously labeling competent student-teachers as incompetent would result in retraining and retesting. Therefore, from a program's short- and long-term efficiency point of view, the number of misclassifications or decision errors should be minimized. Given such serious consequences of decision errors, particularly false acceptance, a cut-off point could be increased high enough such that the only decision errors made concern masters, i.e., all non-masters are correctly rejected.

Should there be a judged difference in seriousness between false acceptance and false rejection, selection of a specific cut-off point relies on reduction of the more serious error. As indicated, shifting cut-off points reduces one error at the expense of the other. Therefore, the direction of shift should be in the direction of the more serious error. For example, if false acceptance is judged twice as serious as false rejection, then the judged seriousness of the error rate would = 2(false acceptance) + 1(false rejection). In other words, the probability of an error should take into account its judged seriousness. Table II presents the correct and error (incorrect) decision rates

with false acceptance being judged equally and twice as serious as false rejection. A cut-off score of 75 would

---

Insert Table II

---

be indicated, particularly for the 2(FA) + 1(FR) situation.

In addition to joint probabilities, conditional or dependent probabilities can be obtained per group or per test score. Joint probabilities are read 'and,' e.g., $P(M_{TS} \cap M_J)$ states the probability that a student is both a master in the true state and also judged a master. Conditional probabilities, however, are read 'given that,' e.g., $P(M_J \mid M_{TS})$ states the probability that a student is judged a master given that he is a master in the true state. It should be noted that the reference group is different for the two types of probabilities. The former is based on the total number of students used, both masters and non-masters, while the latter is based solely on the subgroup of masters.

With the distributions presented in Figure 1, it is possible to obtain a series of conditional probabilities. For example, Table III lists the conditional probabilities of non-masters and masters relevant at each given test score. One would expect a decrease in the probability of

---

Insert Table III

---

non-master with increasing test scores; and, likewise, one would expect an increase in the probability of master with increasing test scores. This is illustrated in Table IV.

Insert Table IV

It should be noted that both Table III & IV deal with the probability of non-master and master at each given test score. Simply, this is the frequency of non-masters at score X and the frequency of masters at score X each divided by the total number of students at score X.

TABLE II

CORRECT & ERROR (INCORRECT) DECISION RATES AT VARIOUS CUT-OFF POINTS AND DIFFERENT DEGREES OF ERROR SERIOUSNESS

CUT-OFF SCORE	CORRECT RATE 1(CA) + 1(CR)	ERROR RATE 1(FA) + 1(FR)	2(FA) + 1(FR)
65	.45 + .30 .75	.20 + .05 .25	.40 + .05 .45
70	.35 + .35 .70	.15 + .15 .30	.30 + .15 .45
75	.30 + .45 .75	.05 + .20 .25	.10 + .20 .30

TABLE III

CONDITIONAL PROBABILITIES OF NON-MASTERS AND MASTERS GIVEN A TEST SCORE

Non-masters		Masters	
P(NM | 30)	decrease ↓	P(M | 30)	increase ↓
P(NM | 40)		P(M | 40)	
P(NM | 50)		P(M | 50)	
P(NM | 60)		P(M | 60)	
P(NM | 70)*		P(M | 70)**	
P(NM | 80)		P(M | 80)	
P(NM | 90)		P(M | 90)	
P(NM |100)		P(M |100)	

*probability of a non-master given test score of 70.

**probability of a master given test score of 70.

TABLE IV

EXAMPLE OF CONDITIONAL PROBABILITIES
AT EACH TEST SCORE

P(NM| 30)	= NA	P(M| 30)	= NA
P(NM| 35)	= NA	P(M| 35)	= NA
P(NM| 40)	= NA	P(M| 40)	= NA
P(NM| 45)	= 100%	P(M| 45)	= 0%
P(NM| 50)	= 100%	P(M| 50)	= 0%
P(NM| 55)	= 100%	P(M| 55)	= 0%
P(NM| 60)	= 95%	P(M| 60)	= 5%
P(NM| 65)	= 80%	P(M| 65)	= 20%
P(NM| 70)*	= 50%	P(M| 70)*	= 50%
P(NM| 75)**	= 20%	P(M| 75)**	= 80%
P(NM| 80)	= 5%	P(M| 80)	= 95%
P(NM| 85)	= 0%	P(M| 85)	= 100%
P(NM| 90)	= 0%	P(M| 90)	= 100%
P(NM| 95)	= 0%	P(M| 95)	= 100%
P(NM|100)	= NA	P(M|100)	= NA

*of the students scoring 70, 50% were non-masters and 50% were masters

**of the students scoring 75, 20% were non-masters and 80% were masters

It is also possible to obtain conditional probabilities of non-masters and masters for a given score and above. This is illustrated in Table V. Calculation, in this case, would

---

Insert Table V

---

consist of the frequency of non-masters at or above score X and the frequency of masters at or above score X each divided by the total number of students at or above score X.

Furthermore, the conditional probabilities of non-masters and masters obtaining scores below X are obtainable. In this case, the frequency of non-masters scoring less than X and the frequency of masters scoring less than X are each divided by the total number of students scoring less than X. For example, $P(NM|<70)$ and $P(M|<70)$ refer to the probability of non-masters and masters, respectively, for students scoring less than 70. Additionally, the conditional probabilities of scoring at or above and below score X can be obtained for non-master and master groups. That is, for each group of non-masters and masters, the frequency of scoring at or above X and the frequency of scoring below X are each divided by the number of students per group. For example, $P(\geq 70| NM)$ and $P(< 70| NM)$ refer to the probability of scoring at or above and below 70, respectively, given that the student is a non-master. And, $P(\geq 70 | M)$ and $P(< 70 | M)$ refer to the probability of scoring at or above and below 70, respectively, given that the student is a master. The researcher, therefore, has a variety of information available to assist in the establishment of cut-off points.

It is possible to reduce the error rate by imposing a grey area of non-decision regarding the overlap of the two frequency distributions; however, this is not practical in an educational setting. It is also possible to reduce both types of error simultaneously by increasing the distance

TABLE V

EXAMPLE OF CONDITIONAL PROBABILITIES AT EACH TEST SCORE AND HIGHER

$P(NM \mid \geq 30)$	= 50%	$P(M \mid \geq 30)$	= 50%
$P(NM \mid \geq 35)$	= 50%	$P(M \mid \geq 35)$	= 50%
$P(NM \mid \geq 40)$	= 50%	$P(M \mid \geq 40)$	= 50%
$P(NM \mid \geq 45)$	= 50%	$P(M \mid \geq 45)$	= 50%
$P(NM \mid \geq 50)$	= 45%	$P(M \mid \geq 50)$	= 55%
$P(NM \mid \geq 55)$	= 40%	$P(M \mid \geq 55)$	= 60%
$P(NM \mid \geq 60)$	= 35%	$P(M \mid \geq 60)$	= 65%
$P(NM \mid \geq 65)$	= 30%	$P(M \mid \geq 65)$	= 70%
$P(NM \mid \geq 70)$*	= 25%	$P(M \mid \geq 70)$*	= 75%
$P(NM \mid \geq 75)$**	= 5%	$P(M \mid \geq 75)$**	= 85%
$P(NM \mid \geq 80)$	= 0%	$P(M \mid \geq 80)$	= 100%
$P(NM \mid \geq 85)$	= 0%	$P(M \mid \geq 85)$	= 100%
$P(NM \mid \geq 90)$	= 0%	$P(M \mid \geq 90)$	= 100%
$P(NM \mid \geq 95)$	= 0%	$P(M \mid \geq 95)$	= 100%
$P(NM \mid \geq 100)$	= 0%	$P(M \mid \geq 100)$	= NA

*of the students scoring 70 or higher, 25% were non-masters and 75% were masters

**of the students scoring 75 or higher, 5% were non-masters and 95% were masters

between the two frequency distributions and/or reducing the variability within each frequency distributions; however, these topics are beyond the present paper.

References

Beggs, D.L. & Lewis, E.L. Measurement and evaluation in the school. Boston: Houghton Mifflin Co., 1975.

Copies of this article are available from: Joseph A. Trzasko, Ph.D.; Department of Psychology; Mercy College; Dobbs Ferry, New York 10522